PEPCAT—A New Tool for Conformational Analysis of Peptides

M. F. O'DONOHUE,[1] E. MINASIAN,[2] S. J. LEACH,[2] A. W. BURGESS,[1] H. R. TREUTLEIN[1]

[1]*Ludwig Institute for Cancer Research and Cooperative Research Center for Cellular Growth Factors, P.O. Royal Melbourne Hospital, Parkville, Victoria 3050, Australia*
[2]*Department of Biochemistry & Molecular Biology, Melbourne University, Parkville, Victoria 3052, Australia*

Received 9 August 1999; accepted 15 November 1999

ABSTRACT: We report a new technique for the efficient analysis and visualization of peptide and protein conformations and conformational relationships, which we have implemented in a computer program called PEPCAT. PEPCAT (an abbreviation for Peptide Conformational Analysis Tool) provides a simple, graphical, and flexible framework that allows the user to define a specific structural feature or juxtaposition of amino acids and to follow the fate of the motif during a molecular dynamics simulation. Here we describe the PEPCAT analysis of the effects of environmental and chemical modifications on conformational preferences of a regulator of hemopoiesis, namely the pentapeptide pyro-EEDCK, and of a conformational transition in the immunosuppressant drug cyclosporin A. PEPCAT, however, can be applied to the conformational analysis of peptides and proteins in general. © 2000 John Wiley & Sons, Inc. J Comput Chem 21: 446–461, 2000

Keywords: cyclosporin A; molecular modeling; transitions; dynamics; conformational search; conformational analysis; hemopoietic peptides

Introduction

Conformations of biological macromolecules are determined by the properties of their molecular building blocks and by their surrounding environment.[1, 2] Macromolecular systems, such as peptides and proteins, often display multiple conformations dependent on temperature and environment. A description and classification of possible conformations is important for the understanding of molecular properties. Methods for the analysis and display of conformational types and energies have been developed for small molecules,[3] but until now simple methods have not been readily available for macromolecules. In this work we describe PEPCAT, a new method for evaluation, comparison of conformational properties, and conformational

Correspondence to: H. R. Treutlein; e-mail: Herbert.Treutlein@ludwig.edu.au

interrelationships, which can be applied to macromolecules.

This article is organized as follows: the remainder of the Introduction section will discuss problems with current conformational analysis tools. The Method section describes our PEPCAT method in detail, and this is followed in the Results section by a description of two examples of a PEPCAT analysis. Finally, in the Discussion section, our method is summarized, and possible extensions are discussed.

CONFORMATION IDENTIFICATION

Some of the crucial questions for peptide structural analysis are how to describe peptide structures and how to determine when they are similar enough that they can be considered to share the same identity or conformation. The most popular method for peptide conformational analysis calculates a similarity measure between each pair of structures, using a root-mean-square difference (RMSD) technique (see below). The structures are then grouped into distinct conformations based on the relative deviation to other observed structures.[4] There are problems with these techniques that can be illustrated by their inability to generate classification hierarchies similar to the three-dimensional fold databases CATH[5] and SCOP[6] without manual inspection as an essential contribution.

TRAJECTORY FILES

Molecular dynamics simulations provide an excellent source of information about the details of pathways between different molecular conformations. The results of molecular dynamics simulations are often saved in a "trajectory file," which contains coordinate sets (or "frames") of the peptide structure saved at regular intervals during the simulation. Each frame consists of the set of all atomic Cartesian coordinates in the peptide at that timestep in the simulation. Because of the large size of trajectory files, and the fact that many separate calculations are needed to ensure statistically relevant results, the data used for conformational analysis can easily occupy large amounts of storage space.

Although automated trajectory analysis techniques exist,[4,7–10] direct viewing of the trajectory file in the form of a playback movie of molecular movement still remains the most common analysis technique for the visualization of conformational changes. However, more elaborate and automated tools are necessary to analyze and represent the more subtle conformational features and changes that occur during a dynamics simulation.

CARTESIAN RMSD MEASUREMENT

Peptide conformational analysis relies heavily on the comparison between structures, and this places a great responsibility on the comparison operator to correctly discriminate between structures. The most popular method to calculate a measure of structural similarity is to overlay the structures using a least-squares fit algorithm, and to compare the Cartesian coordinates of the backbone C^{α} atoms only. The similarity can be expressed in terms of an average deviation of C^{α} atoms from their average position, the root mean square deviation (RMSD).[4]

The RMSD value often gives a good measure of similarity, and is insensitive to small changes in atomic Cartesian coordinates. However, the change of a single backbone dihedral value can lead to a large overall RMSD value. Consider, for example, a backbone dihedral angle change in a residue in the middle of a peptide with an extended β-strand conformation. Under these circumstances, the single RMSD value correctly identifies the quite significant change in the shape of the peptide and the vast number of atoms that have suffered dramatic changes in their Cartesian coordinates. However, it gives no clue that a comparatively simple structural change (rotation of a single dihedral bond angle) is responsible for the difference, and that the two substructures on either side of the changed residue are still identical to the original peptide structure.

Calculation of the RMSD value requires the computation of the optimal overlay of the two peptide structures. This is computationally expensive when compared to other equivalent low level operations such as comparing two numerical values. The high costs of the Cartesian RMSD operation make it an undesirable candidate for comparing large numbers of structures.

DIHEDRAL RMSD MEASUREMENT

A peptide structure can also be represented as a sequence of dihedral bond angle values rather than as a set of Cartesian coordinates. This "internal" coordinate system better represents the local conformations and the degrees of movements that are available to the molecule. Bond length and angle stretching are ignored, as they usually does not significantly contribute to the overall peptide conformation. A backbone dihedral RMSD value can then be calculated between two structures using the

difference in peptide backbone dihedral angle values. This has the advantage of not requiring the expensive overlay operation necessary for Cartesian RMSD comparisons. Structures that differ in only one dihedral angle (such as our previously mentioned example) will, when compared to the original structure, generate a relatively small backbone dihedral RMSD value.

RMSD AVERAGING EFFECT

The RMSD comparison measures the average difference over all coordinate values (this applies in both Cartesian and backbone dihedral angle comparisons). Unfortunately, this destroys the ability to distinguish between large changes in a small number of coordinates, which can be observed in some conformational transitions, from small changes in a large number of coordinate values.

This problem is often accentuated in molecular dynamics simulations (particularly those at elevated temperatures) where the thermal motion induces all coordinate values to undergo small fluctuations and produce a large number of structural variations of each peptide conformation. A structural change, for example, the radical change of a single amide peptide dihedral from *cis* to a *trans*, may go unnoticed against the background structural fluctuations. This averaging effect obviously reduces the effectiveness of an RMSD-based classification scheme to discriminate between groups of structures based on large changes in a small number of residues.

GROUPING STRUCTURES USING RMSD

One popular method for grouping related peptide structures used in the analysis of trajectory files is the calculation of a difference measure between all possible pairs of structures. A clustering algorithm can then sort the structures into a hierarchy of groups and subgroups. A cutoff value can be used to partition the structures into conformational families.[4] This clustering method does not scale up well for the processing of large volumes of data because the number of RMSD comparison operations required increases quadratically with increasing numbers of structures.

This restricts the application of this type of analysis to conformational data sets with hundreds or thousands of frames. However, current molecular dynamics calculations often produce much larger conformational data sets where the use of RMSD-based techniques become computationally prohibitive.

It is also difficult to extend an already concluded analysis with the current method because the addition of new conformational data to an existing set of previously analyzed data set requires the reassessment of all existing data.

Much work has been done in applying clustering techniques to dynamics simulations, and the method is a subject of current research interest.[11–15]

GROUPING STRUCTURES USING A CLASSIFICATION SCHEME

An alternative method for grouping peptide structures is the use of a classification scheme in which an abstract set of rules is applied to identify similar peptide structures. A commonly used classification scheme is a residue–residue contact map.[16, 17] Two residues are considered to be in contact if any atom of one residue comes within a predetermined cutoff radius of an atom of the other residue. Different conformations of peptide structures will have different patterns of residue–residue contacts. This classification scheme does not use a continuous solution space, as the peptide Cartesian coordinates does, but instead, maps these structures into a discrete solution space. This new coordinate system has a number of advantages. A conformation can be simply represented as a $K \times K$ matrix, where K is the number of residues in the peptide, with cell values containing a Boolean true or false value for contact between residues. The size of the solution space, which is infinite in atomic Cartesian coordinate space and is still infinite (but with less dimensions) in dihedral angle space, is reduced to $2^{K \times K}$ possible contact matrices. A comparison between two different contact maps can be calculated easily by comparing the two matrices. And finally, classification is computationally less expensive, as it does not require a comparison between all possible pairs of structures, but only a calculation performed once on each structure. The identification of a discrete conformational "state" in this manner opens the way for use of traditional state space search and analysis techniques commonly used in artificial intelligence applications.[18] Contact map representations have been used, for example, to analyze results from the series of CASP experimental prediction trials.[19]

More general classification schemes not based on a contact matrix, but on the application of a number of independent descriptors can be used to identify a conformation. Lambert et al.[20] have used descriptors based on regions in ϕ–ψ maps to analyze and predict the conformation of small peptide

sequences. The ϕ–ψ map for a peptide was divided into four regions, and a peptide conformation was classified as the sequence of residue descriptor values. Bravi et al.[21] used a mixture of distance and dihedral descriptor measurements, each with an individual cutoff comparison threshold value to develop a procedure to define and compare molecular conformations. A nonhierarchical clustering technique is then employed to group the observed conformations.

SUBJECTIVE NATURE OF CLASSIFICATION

The definition of a peptide conformation depends upon the needs of the investigator and the question being asked. What constitutes a conformation in a ligand docking problem may differ substantially from that required for evaluating protein flexibility. This subjective nature of classification makes it difficult to devise a single structural classification scheme. However, it does allow for the development of a "tool kit" that can be used to define the structural criteria that will allow the researcher to discriminate between different protein or peptide conformations. For instance, in a ligand-docking problem, only residues near the binding pocket need to be considered in detail; changes in other residues can either be ignored or monitored at a lower level of accuracy.

It is worth noting that grouping of structures using only the three dimensional coordinates as the basis for similarity, may also be misleading in some cases. Two conformations, which may be clustered together based on their structural similarity, may have significantly different potential energies, and/or be separated by great energy barriers on the potential energy surface of the molecule.

THE PEPCAT METHOD

To address many of the problems encountered in conventional conformational analysis methods we have developed a method that allows a flexible definition of conformational states and an efficient comparison of a large set of structures. Our method is called PEPCAT (for Peptide Conformational Analysis Tool), and consists of five steps: conformation definition, classification, comparison, trajectory analysis, and visualization of the results. Emphasis was also put on an intuitive display of the results, which facilitates the recognition of important conformational features present in molecular dynamics trajectory files. The PEPCAT methodology is first defined in the next section. This is followed by a detailed description of two applications of PEPCAT in the Results section.

Methods

There are five steps in the PEPCAT analysis technique: first, the user defines the geometric properties of the peptide relevant to the particular problem; second, structures are processed using the classification scheme; third, the similarities between structures are calculated using a generalized distance measure; forth, trajectory files are reexpressed as a compact sequence of conformational states rather than coordinates; and the last step uses the new trajectory files and a database of the known conformations built during the previous steps to analyze as well as visualize conformational properties and conformational transitions. Each step is described in detail below. The terminology specific to the PEPCAT methodology is italicized in the text, and is described in Table I. A visual guide to the PEPCAT methodology is also shown in Figure 1.

CONFORMATION DEFINITION

A classification scheme is devised by the user to identify conformations of a peptide, and is coded into an input file for the PEPCAT program.

A *descriptor*, such as the ϕ–ψ values for a particular residue can be selected from the supported *descriptor types* for monitoring in the peptide. A map is constructed for each *descriptor* to map all values of the generalized peptide property to a discrete numeric *descriptor state*. This allows the user to specify the level of accuracy for monitoring a particular descriptor. PEPCAT currently supports three *descriptor types*: ϕ–ψ regions (see Fig. 2a), angle range values (see Fig. 2b), and atomic distances (see Fig. 2c). Each *descriptor* maps a generalized peptide property to a discrete numeric *descriptor state*.

A set of *descriptors*, called a *classification scheme,* is chosen to distinguish different conformations of the peptide. Each distinct set of *descriptor values* resulting from the application of the *classification scheme* to a peptide structure is defined as a unique peptide conformation or *conformational state*.

CLASSIFICATION

The PEPCAT program applies the constructed classification scheme to peptide structures resulting in a set of *descriptor values* that identify each input structure. PEPCAT also uses this information to

TABLE I.
PEPCAT Terminology.

Descriptor type	A generic type of measurement such as an angle, distance or ϕ–ψ map value that can be made of a peptide property.
Descriptor	A specific property of a peptide to be monitored. The precise residue or angle to be monitored is identified, the descriptor type, and an appropriate map is designated to map all values of the descriptor type to a discrete numerical value.
Descriptor value or descriptor state	The discrete numerical value returned from the application of a specific descriptor to a peptide structure.
Classification scheme	An ordered set of descriptors that are to be applied to a peptide structure to discriminate between different conformations.
Descriptor value set	An ordered set of descriptor values containing the results, of application of each descriptor described in the classification scheme to a peptide structure.
Conformational state	A peptide conformational state is identified by a distinct set of descriptor values.
State identification number or state id	A unique numeric identifier assigned to each conformational state. Each new state observed is given a unique numeric identifier.
Known states database	A database that contains a set of previously observed set of conformations. Each identified by a unique state id, descriptor value set, and containing a representative structure for each chemical system investigated.
Conformational difference	A measure of the difference between two conformational states.
Chemical system	A distinct environment or peptide chemical modification used in the analysis. Many different chemical systems can share the same set of known states database.

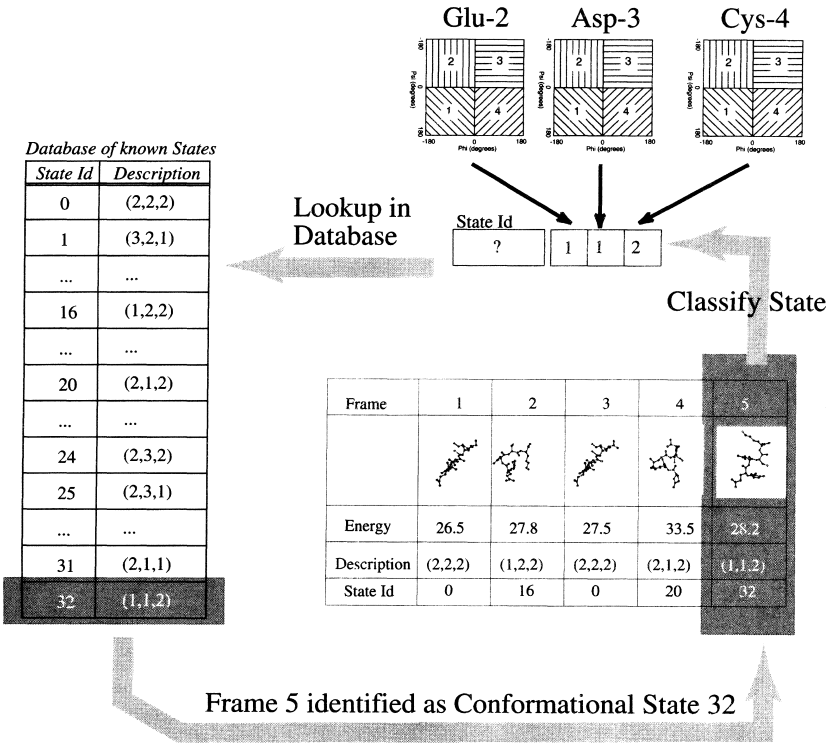

FIGURE 1. Methodology: this diagram shows the classification of a pyro-EEDCK trajectory *in vacuo* using the PEPCAT methodology. A trajectory consists of a time sequenced set of peptide coordinates (bottom right). Each frame is classified using a set of user specified descriptors (top right), and the set of values is looked up in a database of known states (left). If the set of descriptor values is not found in the database, a new entry is created. The conformational state identifier is returned (left) and stored with the frame. See text for more detail.

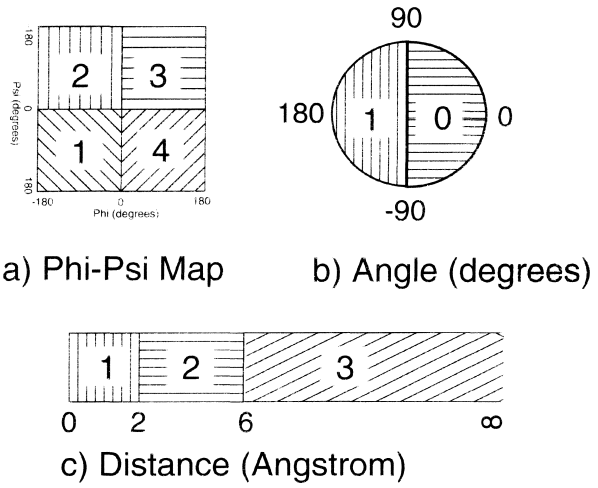

FIGURE 2. Descriptor types: the currently supported descriptor types are: (a) a ϕ–ψ map region; (b) a dihedral angle; and (c) a distance between two atoms. Each of the descriptors map every possible value to a discrete, numerically identified state. The number of states and map structure are not limited to the ranges shown here, and can be altered by the user. The examples shown here can be used to: distinguish between four ϕ–ψ conformations of a residue (a); identify the *cis* or *trans* state of a peptide bond (b); and identify contact, close or no contact between two atoms (c).

maintain a database of all observed peptide conformations.

The known conformation database is initially empty, but as new *conformational states* are identified in the classification procedure, entries are appended to the database to describe these new states. Each new state is assigned a unique state identification tag (*state id*), and a new database entry is written. A *conformational state* database entry contains the unique *state id*, the distinct *descriptor value set* describing the state, and the coordinates of a reference peptide structure occupying the state.

As PEPCAT classifies each peptide structure using the *classification scheme* the resulting *descriptor value set* is searched for in the database of known conformations. If the *descriptor value set* (i.e., a particular conformation) does not already exist in the database, a new *state id* is generated and a new database entry is written, with the *state id*, the defining *descriptor value set*, and the coordinates of the current structure, stored as the reference structure. The new *state id* is then returned from the classification procedure. Otherwise, if the *descriptor value set* was found in the database then the database entry matching the found *descriptor value set* is examined. If the energy value of the current structure is less than the energy value of the reference structure, then the reference structure in the database is replaced by the current structure. The *state id* of the found database entry is then returned from the classification procedure.

The comparison of the energy values ensures that the reference structure for each state has the lowest energy for all structures in this state, that have been examined to date.

As peptide structures are classified under a variety of conditions, for example, different environments, etc., the data in the known conformation database is held at two levels. The first level identifies the *conformational state*, and contains the *state id* and the *descriptor values set* that describes the state. At the second level, each *chemical system* has stored its own separate and distinct minimum energy value and representative structure for each of the known conformational states. In this way the *conformational state* identifiers and descriptors can be shared among any number of investigated *chemical systems* and comparisons between different *chemical systems* are simplified.

For each chemical system a database of minimized conformational structures is maintained in addition to the database of reference structure, and a link is maintained from the reference structure to the minimized reference structure. After classifying a large number of structures, the database of known states is subjected to a clean-up step. Every new or changed *chemical system* reference structure is subjected to an energy minimization, and the database of minimized conformations is updated. These minimized structures are then reevaluated using the classification procedure, and the resulting *state id* is stored with the minimized reference structure. This reprocessing of reference structures helps ensure that low-energy values are used for the reference structures. Sometimes a high energy *conformational state* will, after minimization and classification, result in a state different from the original structure. Where this occurs, the stored link from the original classification to the new classification can be used to identify the new state achieved upon minimization of the reference structure. In this way both the original state and the minimized state are available for reference in further analysis.

COMPARISON

The *conformational difference* Δ between two states A and B is determined by the differences in their

descriptor value sets and is calculated as follows:

$$\Delta = \sqrt{\sum_{i=1}^{n} \delta\big(A[i], B[i]\big)^2} \qquad (1)$$

where

$$\delta(a,b) = \begin{cases} 0, & \text{if } a = b \\ 1, & \text{else.} \end{cases}$$

where $A = (A[1], \ldots, A[n])$, $B = (B[1], \ldots, B[n])$ are descriptor value sets; $A[i]$, $B[i]$ are descriptor values of the descriptor i; and n is the number of descriptors in the *classification scheme*. The square of the result of the comparison function δ is actually redundant for the current choice in eq. (1), but is relevant if modifications are made to the function δ.

TRAJECTORY ANALYSIS

PEPCAT processes molecular dynamics trajectories to produce a new much smaller trajectory file. Each frame of the original trajectory is classified using the classification procedure, and an entry is appended to the new trajectory file containing two identifiers the *state id*, and the original frame energy value.

VISUALIZATION

The new trajectory file format can easily be subjected to further analysis techniques. Two reports are currently available within PEPCAT.

The conformational population distribution report (see Fig. 3) displays the percentage of the trajectory frames occupying each *conformation state*. The *conformational state* value of each frame in one or more new format trajectories is counted, and the distribution of occupation of each conformation in the database of known states is displayed in bar graph form.

The conformation transition map (Figs. 4, 5, and 9) is a visual map of the conformational changes that occur during the course of a dynamics trajectory. The new PEPCAT formatted trajectories contain a time sequenced set of *states id* numbers and their energies. These are used to build a directed graph of conformational states, which connects two conformational states between which a transition was observed. This graph is then mapped onto a two-dimensional surface.[22] The resulting layout provides a simple picture of the often complex relationships between the different *conformational states*.

COMPUTATIONAL DEMANDS

The CPU time demands for the classification of a single frame depends upon the number of descriptors that are being monitored. A typical analysis requires one descriptor per protein residue, and is constant for any given classification scheme. The time taken to classify a single frame does increases linearly with the length of the peptide being analyzed. However, even for proteins of several hundred residues length, classification takes less than a second of CPU time, and this is still much less than the time required to perform the overlay operation required by RMSD-based techniques.

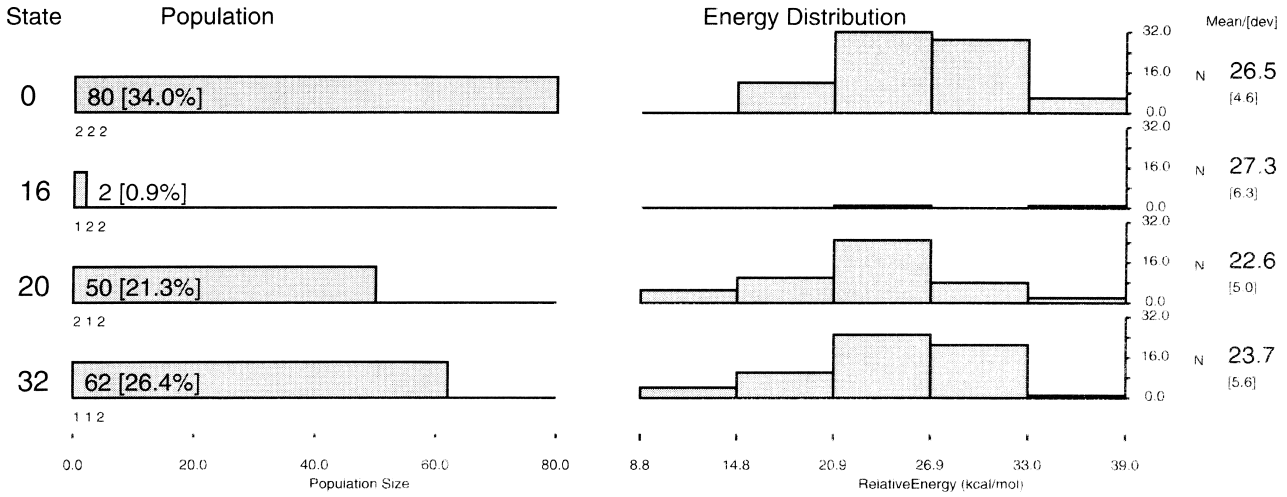

FIGURE 3. Conformation distribution: the occupancy of all conformational states for the trajectory can be calculated (although only four are shown). The report displays for each state, the number of frames occupying each conformation, as a percentage of the total trajectory frames (on left) and in a bar graph form (on right). The mean energy and standard deviation for each state is printed along with a small graphical display of the energy distribution (on right).

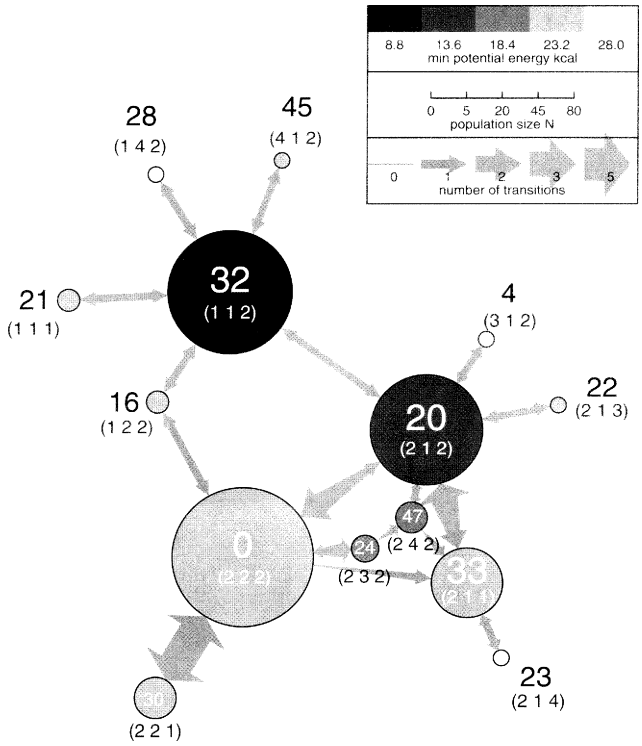

FIGURE 4. Pyro-EEDCK *in vacuo* transition map: a conformational transition map analysis for a simulated molecular dynamics trajectory of pyro-EEDCK performed in vacuum. Each state visited during the trajectory is represented by a circle, with an area proportional to the number of occurrences of the state in the trajectory. The state is identified by its state id, in bold type, and in a slightly smaller font, the defining set of descriptor values is also shown. The state's gray color shade identifies the minimum energy value. The lines between conformations show the conformational transitions that occurred in the trajectory. The arrowhead identifies the direction of the transition and its width proportional to the number of transitions that occurred.

The use of a more finely graded grid size for the descriptors has no effect on the CPU time required. Because the time taken to classify a single frame is constant, the time required analyzing a dynamics trajectory grows linearly with the number of frames in the trajectory.

However, although there is no detrimental effect on CPU time, the use of finer graded grids and the analysis of larger molecules with more descriptors does have an exponential effect on the potential size of the known states database. The exact choice of descriptor grid size and simulation temperature are important factors in determining the number of states visited in molecular dynamics trajectories and, consequently, the size of the known states database. The size of the database is important in determining the CPU times required for some of the analysis procedures.

The CPU time required to produce the displays generally has two components: the generation of a comparison matrix, which grows quadratically with the number of distinct states (which is much less than the number of trajectory frames, and hopefully, much less than the potential number of states), and the creation of the displays and reports, which grow linearly with the number of frames analyzed.

For example, using a Pentium II 400 MHz running the Linux operating system on data sets that have less than 10 descriptors and less than 1000 states (i.e., data sets of a similar size to those used in this article), trajectories of 200 frames are classified in minutes, and the analysis procedures take less than 10 min CPU time. Increasing the number of descriptors to 2000 increased the CPU time to process a 200 frame trajectory to approximately 10 min. The time taken to perform the analysis routines is independent of the number of descriptors, but is dependent upon the size of the database of known states. The largest database we have built so far contains 7300 states, and the calculation of a combined transition map for 336,000 frames in a trajectory takes approximately half an hour of CPU time.

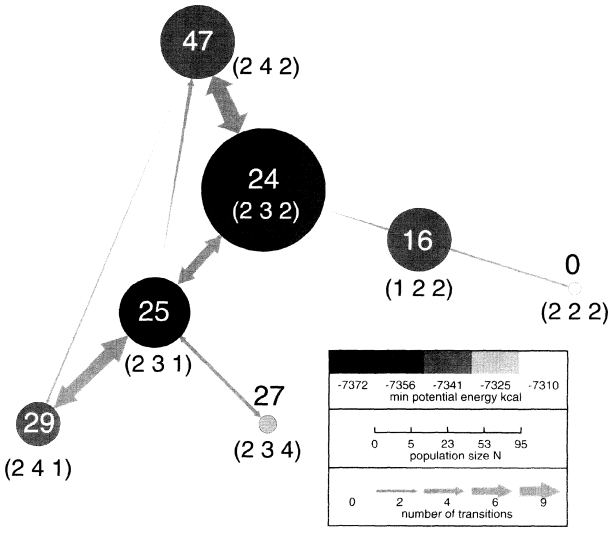

FIGURE 5. Pyro-EEDCK in solution transition map: a conformational transition map analysis for a simulated molecular dynamics trajectory of pyro-EEDCK performed in explicit simulated water. See Figure 4 to compare with a similar analysis performed for an *in vacuo* simulation, and also see the Figure 4 caption for a description of the diagram components.

Results

In this section the use of the PEPCAT method is demonstrated with two examples: an analysis of the conformational space for pyro-EEDCK, and a conformational transition in cyclosporin A. Where PEPCAT specific terminology has been used, it has been italicized in the text and is described in Table I. For a description of these terms and further details of the PEPCAT method, please consult the methods section of this article.

PEPCAT ANALYSIS OF PENTAPEPTIDE PYRO-EEDCK

The pyro-EEDCK monomer inhibits hematopoietic stem cell proliferation of colony stimulation units in the S state (CFU-S) by maintaining them in a quiescent state while they are exposed to the radiation or cytotoxic drugs used during cancer therapy. The acute and chronic bone marrow toxicities are the major limiting factors in oncology. The acute effects include neutropenia and thrombopenia by increasing patient susceptibility to infections and hemorrhage. They interfere with the treatment and limit its therapeutic effects.[23–25] Pyro-EEDCK appears to be a candidate drug for cancer treatment having a physiologic role for the protection of the hematopoietic system. Conservative changes to the chemical sequence as in: EEDCK, pyro-EDDCK, pyro-EEDMK, and pyro-EEDSK, reduce or destroy the inhibitory properties of this peptide.[24] Pyro-EEDCK-like sequences have been identified in the $G_{i\alpha}$ chains of GTP-binding proteins at position 63–67, proximal to the major phosphorylation site.[24, 26] Oxidation of the cysteine thiol groups of pyro-EEDCK can lead to formation of a disulphide-bridged homodimer $(\text{pyro-EEDCK})_2$ which, unlike the monomer, is a potent stimulator of hematopoiesis.[27]

Molecular Dynamics Calculations

A number of molecular dynamics simulations of pyro-EEDCK were performed. A 400-ps simulation was carried out at a temperature of 300 K *in vacuo*, with a dielectric constant of 10 using the program Discover (Molecular Simulations Inc., San Diego, CA). The CFF91 force field[28] was used with a simulation timestep of 1 fs and interatomic interactions were reduced to zero using a switching algorithm for atom pair distances between 8.5 and 13.0 Å and interactions over 13 Å were ignored. At 2-ps intervals the coordinates were written to a trajectory file. Additional dynamics simulations of pyro-EEDCK were calculated at a variety of temperatures from 300 to 1000 K, for time periods ranging from 200 ps to 2 ns. These calculations used the same protocol except that at 2-ps intervals a 300-cycle energy minimization was performed before the frame was written to the trajectory file.

Molecular dynamics simulations of pyro-EEDCK were also performed in explicit water. A single molecule of pyro-EEDCK was placed at the center of a cubic box with side length of 27 Å, which was then filled with CFF91[28] (Molecular Simulations Inc.) water molecules using the Insight II package (Molecular Simulations Inc.). Simulations were then performed using the Discover program as with the *in vacuo* calculations described above, except that a dielectric of 1 was used and periodic boundary conditions were applied.

Using the same set of protocols, the conformational properties of peptides closely related to pyro-EEDCK were also investigated using molecular dynamics in both *in vacuo* and solvated environments. Modifications included simple residue substitutions and deletions, namely EEDCK, pyro-EDDCK, pyro-EEDMK, pyro-EEDSK, and extended peptides SEEDCKN and YSEEDCKNY as found in the $G_{i\alpha}$ protein sequence.

Classification

Classification of the pyro-EEDCK and related peptides makes use of the ϕ–ψ map *descriptor type* (see Fig. 2a). The *descriptor* used maps the ϕ–ψ region into four different numerical values, corresponding to the four quadrants of the ϕ–ψ plane. This partition scheme was suggested by an analysis of the distribution of ϕ–ψ dihedrals during initial calculations.

The *classification scheme* for identifying pyro-EEDCK conformations (see Fig. 1) is based on ϕ–ψ map regions for the three central residues of pyro-EEDCK, namely Glu2, Asp3, and Cys4. These three residues form the *descriptor set* that defines the classification scheme. For example, the notation, or descriptor set (2,3,1) would describe a conformation where the Glu2 (ϕ, ψ) conformation would be found in the upper left quadrant of the Ramachandran plot, Asp3 in the upper right and Cys4 in the lower left quadrant.

Analysis

To illustrate the main steps of a PEPCAT analysis, we will discuss the analysis of a pyro-EEDCK dynamics simulation trajectory data file in detail. We assume that other trajectories for pyro-EEDCK have been processed previously, and a database has been populated with the states found in these trajectories. Before processing this particular dynamics trajectory the database is loaded (see table on the left of Fig. 1). Each entry in the known states database contains the states *identification number (state id)*, with their classification *descriptor set* and a representative coordinate structure. The input trajectory is then processed one frame at a time (bottom right of Fig. 1). In Figure 1 the processing of frame 5 from the trajectory is depicted. Prior to this, frames 1 to 4 will have already been processed. The fifth frame in the trajectory (see bottom right of Fig. 1) is categorized according to the classification scheme (see top right of Fig. 1), which results in a *descriptor set* of (1,1,2) (see center right of Fig. 1). This set (1,1,2) is then used to search through the database of known states and it is identified as *conformational state* 32 (see bottom left of Fig. 1). Frame 5 has, therefore, been identified as *conformational state* "32." Because the state already exists, the peptide structure in frame 5 is compared to the reference structure for the state 32, and replaces the database reference structure for state 32 if it has a lower energy value. This completes the processing of frame 5.

Alternatively if we assume that state 32:(1,1,2) does not yet exist in the database and the known states database only contains states 0–31 [which would occur if this is the first time that the (1,1,2) structure has been observed], then the database search using the frame 5 *descriptor set* (1,1,2) would not have found a match. Thus, a new pyro-EEDCK conformational state would have been identified and new database entry would be created to describe this new conformation. The new entry would be assigned the next *state id* number (32), have the *descriptor set* (1,1,2) and a reference structure taken from frame 5. The entry would be appended to the database of known states effectively creating the state 32 as observed in Figure 1.

Processing then continues through the rest of the frames in the trajectory (not shown here). At the completion of a trajectory analysis the updated database of known pyro-EEDCK states is saved and a simplified trajectory containing only the *state id* and energy value for each frame are written to a file. The database of known states is then subjected to a minimization clean-up step (see the Methods section). The cleaned up database and the new trajectory file can now be used as input for the standard PEPCAT reporting routines. Figure 3 shows the conformational distribution report that was produced for all observed states; however, only the four most populous states are shown here.

Comparison

The comparison procedure is described in the Methods section above, and makes use of a mathematical function $\delta(x, y)$ which yields 0, if $x = y$ and otherwise 1. For example, the "conformational difference" Δ between state 0:(2,2,2) $= (a_1, b_1, c_1)$ and state 1:(2,1,2) $= (a_2, b_2, c_2)$ is calculated as follows:

$$\begin{aligned}\Delta &= \sqrt{\delta(a_1, a_2)^2 + \delta(b_1, b_2)^2 + \delta(c_1, c_2)^2} \\ &= \sqrt{0 + 1 + 0} \\ &= 1 \end{aligned} \quad (2)$$

Similarly, for the states 0:(2,2,2) and 5:(1,1,2) $\Delta = 2$, quantifying the concept that states 0 and 1 are more similar to each other than states 0 and 5. A table can be built up by applying the comparison procedure to all the known states, and these data can be used for further analyses.

Results

The analyzed trajectory file used in our example of pyro-EEDCK *in vacuo* at 300 K contained 14 distinct pyro-EEDCK conformational states out of the 64 possible descriptor set conformations. State 0,

with an extended beta structure (2,2,2), was by far the most preferred conformation (34%), and this state, together with the next three most populous states 20:(2,1,2), 32:(1,1,2), and 33:(2,1,1), accounted for almost 90% of the frames in this trajectory (see Fig. 3).

PEPCAT can also be used to visualize the pathways traversed between the conformations observed during the dynamics simulation. Figure 4 shows such a transition map display. The area of each circle represents the frequency of occupation of the state. The large circles representing the states 0, 20, 32, and 33 easily identify them as the four major conformations. Transitions between states are shown as arrows where the width of the arrow head is proportional to the number of transitions observed. For example, there is a high fluctuation between states 0 and 30. Both states have a similar energy, as indicated by the color of the circles. States with lower energy are colored dark gray, states with higher energy are shown in light gray. Not surprisingly, states with highest energy (white) are not visited very often. Some of the remaining 11 states reached during the molecular dynamics are intermediates on a pathway between two major states (see states 16, 24, and 47 in Fig. 4). Other states are higher energy variants of one of the major states, which were occupied for brief periods of time before returning to one of the major states (see states 4, 45, 28, etc., in Fig. 4).

When simulations were performed *in vacuo* at increasingly elevated temperatures the populations of the major states declined as the temperature increased, and small populations of additional conformational states were observed. Eventually, at 1000 K transient populations of all 64 possible states were observed during 20-ns simulation runs. Even at these greatly increased temperatures the low-energy conformations preferred at lower temperatures were still among the most populated (Table II).

Modification of the sequence or side chains and extension of the N- or C-terminus did not significantly alter the conformational preferences of the central residues from those preferred for the wild-type pyro-EEDCK molecule. This was also true with simulations at a variety of temperatures between 300 and1000 K.

When the simulations were performed in a box of water, with periodic boundaries, there was a dramatic effect on the conformational preferences of pyro-EEDCK (see Fig. 5). A new preferred conformational state, 24:(2,3,2) was observed and this occupied 47% of the frames in the trajectory. Again, a large percentage (91.5%) of the trajectory was occupied by the top four states, 24:(2,3,2), 47:(2,4,2), 25:(2,3,1), and 16:(1,2,2) (Table III), but these were different conformational states from those preferred in the previous *in vacuo* calculations (see Fig. 6). Visual inspections of the conformations preferred in water (see bottom of Fig. 6) shows an orientation that maximizes the exposure of the significantly charged atoms found in the backbone side chain to the solvent water, with a small cluster formed by the weakly charged atoms in the pyro-Glu1 and Cys4 residues.

TABLE II.
The Four Most Populated States for Vacuum Simulations of pyro-EEDCK at Different Temperatures.

State	Description	Occupancy (%) At 300 K	Occupancy (%) At 1000 K
0	(2,2,2)	34	13.7
32	(1,1,2)	26.4	4.4
20	(2,1,2)	21.3	8.0
33	(2,1,1)	8.5	3.8

PEPCAT ANALYSIS OF CYCLOSPORIN

Cyclosporin A (CsA) is an 11-residue cyclic peptide used as an important therapeutic agent for the prevention of graft rejection in clinical organ transplantation. Its amino acid sequence MeBmt-Abu-MeGly-MeLeu-Val-MeLeu-Ala-(D-Ala)-MeLeu-MeLeu-MeVal, contains two uncommon amino acids: (4R)-4-((E)-2butenyl)-4, *N*-dimethyl-L-threonine (Bmt) and L-α-aminobutyric acid (Abu). It has several *N*-methylated residues (residues 1, 3, 4, 6, 9, 10, and 11)[29] indicated by the "Me" prefix in the sequence above. CsA binds to a receptor cyclophilin (Cp) and induces an interaction between the CsA–Cp complex and the phosphatase calcineurin.[30] The

TABLE III.
The Four Most Populated States for Simulation of pyro-EEDCK with Explicit Water.

State	Description	Occupancy (%)
24	(2,3,2)	47
47	(2,4,2)	17
25	(2,3,1)	15
16	(1,2,2)	12.5

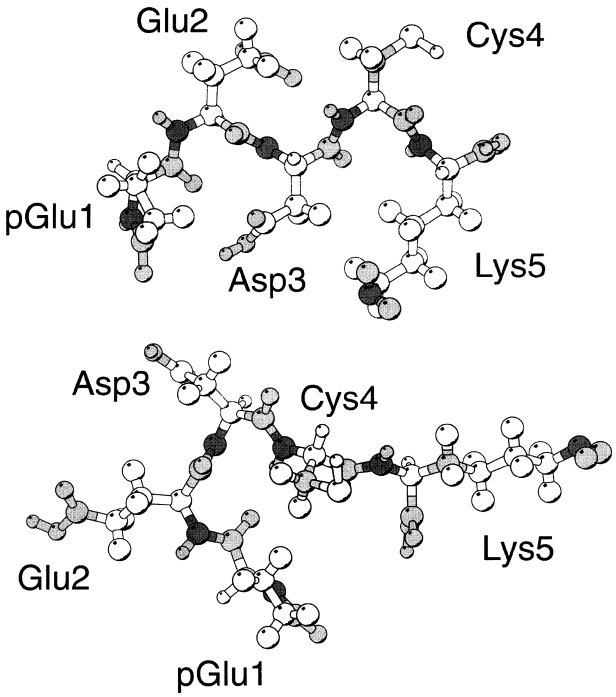

FIGURE 6. pyro-EEDCK conformations: two pyro-EEDCK conformations with atoms shaded according to their atomic partial charge between −0.5 and +0.5. The top figure is the reference structure for state 0:(2,2,2) which forms an extended structure and is the preferred conformation (with 37%) in the *in vacuo* simulations. The bottom conformation is the reference structure for state 24:(2,3,2), which is the preferred conformation for the peptide in water. It adopts a conformation that maximizes the exposure of the significantly charged atoms of both the backbone and side chains to the solvent.

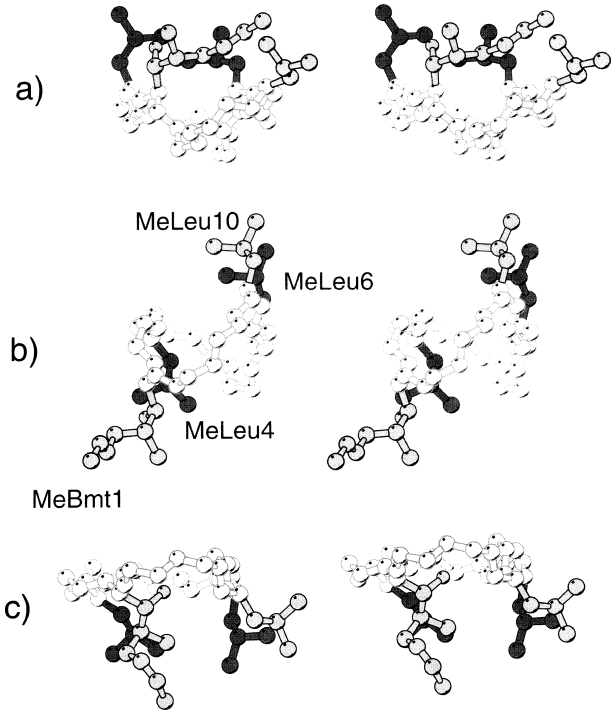

FIGURE 7. CsA conformations: conformations of CsA found in the dynamics calculation. From top to bottom, (a) the unbound ("free") conformation of CsA, (b) one intermediate conformation observed in a molecular dynamics simulation, and (c) the "bound" conformation of CsA. These conformations have been classified (see the text for details) as states 0:(1,1,1,1,0,1), 29:(2,1,0,2,0,0), and 36:(2,2,2,2,0,0), respectively.

complex inhibits the signal transduction pathways that lead to T lymphocyte activation.[31]

The conformation of unbound ("free") CsA in a nonpolar solvent (Fig. 7a) varies considerably from the conformation of CsA that is bound to its receptor Cyclophilin (Fig. 7c). The dominant hydrophobic cluster formed by the side chains of four residues MeBmt-1, MeLeu-4, MeLeu-6, and MeLeu-10 is now located on the opposite side of the molecular ring plane. In addition, the peptide bond at residue Leu9 changes from *cis* in the free conformation to *trans* in the bound conformation. In short, the molecule has been turned inside out.[32]

Molecular Dynamics Calculations

The coordinates of free CsA in CCl_4 solution were obtained from the NMR structure deposited in the Cambridge Crystallographic Databank[33] (access code: DEKSAN[34]). The coordinates for the bound conformation were obtained from the X-ray structure of the CsA-Cp complex[35, 36] (Brookhaven Protein Databank[37] access code: 1CWA).

The molecular dynamics trajectory that is used as the basis for the PEPCAT analysis described here was reported by O'Donohue et al.[38] for an investigation of conformational changes in CsA. The trajectory monitors the transition from a free CsA conformation *in vacuo* at 600 K through a number of conformational changes to a structure nearly identical to the bound conformation (backbone RMS deviation of 0.53 Å). The simulation was performed using the program XPLOR[10] with the CHARMM22 force field.[39]

Classification

The classification scheme employed here for CsA conformational analysis makes use of an angle *descriptor type* (see Fig. 2b) to monitor the angle that a side chain makes relative to the plane defined by

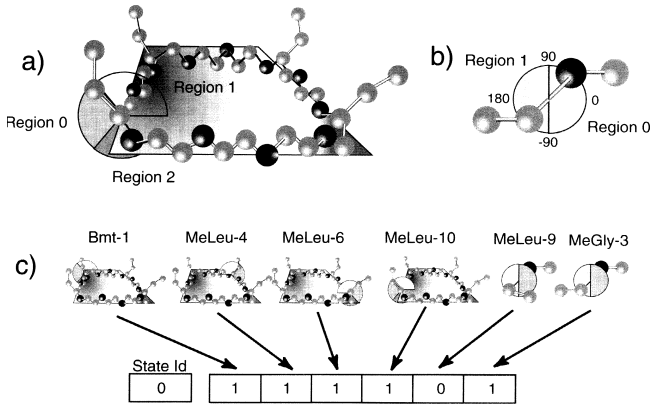

FIGURE 8. CsA classification: the PEPCAT classification scheme used for CsA monitors the orientation of the sidechain relative to the backbone ring plane of the four resides that form the dominant hydrophobic cluster. The two peptide bonds MeGly-3 and MeLeu-9 are also monitored, as they are known to have been observed in both *cis* and *trans* orientations. See text for details.

the ring of backbone atoms in the cyclic molecule. This angle is monitored for all four side chains of the dominant hydrophobic cluster (see above), and is approximated by a dihedral angle defined by the following four atoms: the C^{β} and C^{α} atom positions of the residue whose side chain we wish to monitor; the C^{α} atom position in the residue of the dominant cluster that is closest in the ring plane; and a C^{α} atom position in a residue of the cluster on the other side of the ring plane. The angle map value is classified into three regions: region 0, parallel to the plane (value between $-30°$ and $+30°$); region 1, above the plane (value $> 30°$); and region 2, below the plane (value $< -30°$) (see Fig. 8a). Four descriptors are required to monitor the relative side chain positions for each residue in the hydrophobic cluster. The detailed definition of the descriptors, i.e., the atoms that make up the angle descriptors are: Descriptor 1: Bmt1 ($C^{\beta}1$, $C^{\alpha}1$, $C^{\alpha}10$, $C^{\alpha}6$); Descriptor 2: MeLeu4 ($C^{\beta}4$, $C^{\alpha}4$, $C^{\alpha}6$, $C^{\alpha}10$); Descriptor 3: MeLeu6 ($C^{\beta}6$, $C^{\alpha}6$, $C^{\alpha}4$, $C^{\alpha}1$); and Descriptor 4: MeLeu10 ($C^{\beta}10$, $C^{\alpha}10$, $C^{\alpha}1$, $C^{\alpha}4$).

In addition to the four side chain orientations, ω-dihedral angles of two peptide bonds are also monitored using a angle *descriptor type*. Here, the angle map value is divided into two regions: region 0 (value between -90 and $+90$) identifies a *cis* conformation and region 1 (value larger than $+90$ or less than -90), which corresponds to the *trans* conformation (see Fig. 8b). The MeLeu-9 peptide bond is monitored, as it differs in the bound and free conformations. The MeGly-3 peptide bond is monitored as it has been found in a *cis* conformation in a CsA analogue.[40] The two peptide bonds descriptors are: Descriptor 5: Gly3 ω dihedral angle ($C^{\alpha}3$, C3, N4, $C^{\alpha}4$); and Descriptor 6: Leu9 ω dihedral angle ($C^{\alpha}9$, C9, N10, $C^{\alpha}10$).

These six descriptors make up the *classification scheme* used for the analysis of CsA conformations (see Fig. 8c).

Results

The free CsA conformation, which corresponds to *conformational state* "0," has a descriptor value set of (1,1,1,1,0,1). This conformation has all four dominant hydrophobic side chains sitting above the molecular ring plane, a *trans* peptide bond for MeGly-3, and a *cis* peptide bond for MeLeu-9. The bound CsA conformation corresponds to state 36:(2,2,2,2,0,0), where all four side chains sit below the backbone ring plane, and MeLeu-9 has a *trans* peptide bond. Figure 9 shows the transition map diagram produced by PEPCAT. We can clearly see transition pathways from the free conformation (state 0) to the bound conformation (state 36), through a highly populated intermediate conformation (state 29).

The lowest energy pathway from the free to the bound conformation of CsA is (see Fig. 9): state 0:(1,1,1,1,0,1)—"free" conformation; state 2:(1,1,1,1, 0,0); state 21:(1,1,1,0,0,0); state 29:(2,1,0,2,0,0)—"intermediate" two side chains flipped and one in the ring plane; state 35:(2,0,2,2,0,0); state 36:(2,2,2, 2,0,0)—"bound" conformation.

The first major change in the free conformation (state 0) is the transition to state 2:(1,1,1,1,0,0) where the MeLeu-9 *cis* peptide bond has been lost. In the move to state 21:(1,1,1,0,0,0), the MeLeu-10 side chain repositions itself parallel to the backbone ring plane. Then the transition to the highly populated

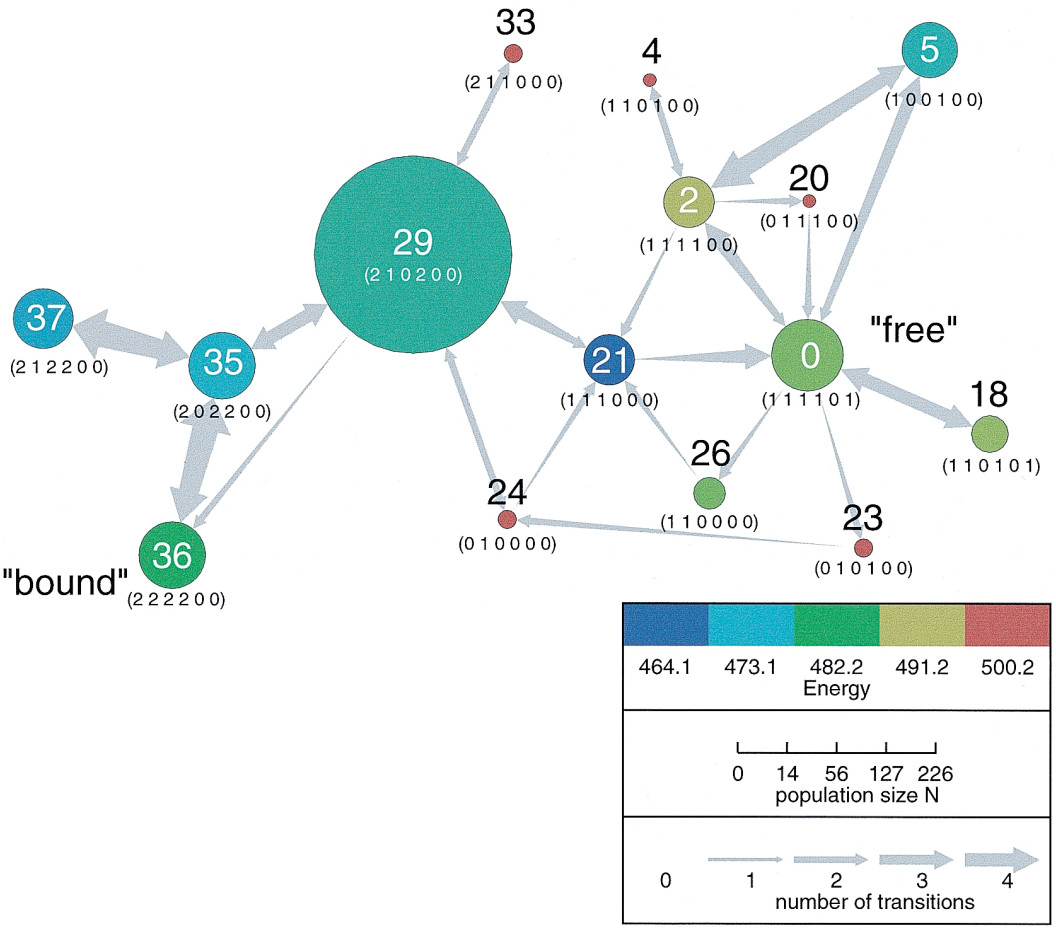

FIGURE 9. CsA conformational transition map: this diagram displays the transitions between the conformations of Cyclosporin A found during the trajectory analysis. Several pathways can be seen from the "free" conformation (state 0) to the "bound" conformation (state 36). A stable intermediate conformation (state 29) was also found (see text for more details).

intermediate state 29:(2,1,0,2,0,0) involves the rotation of both MeBmt-1 and MeLeu-10 side chains around the outside of the ring plane to a position below the backbone ring plane and the MeLeu-6 side chain moving to an orientation parallel to the ring plane. In the move to state 35:(2,0,2,2,0,0), the MeLeu-6 side chain now occupies a position below the ring plane, and the LEU-4 side chain is now parallel to the ring plane. In the final transition to state 36:(2,2,2,2,0,0) the MeLeu-4 side chain joins the other three dominant hydrophobic side chains below the backbone ring plane.

An alternative higher energy pathway for the transition from state 0 to the highly populated state 29 can also be observed (see Fig. 9) via state 23:(0,1,0,1,0,0), where the *cis* peptide bond is lost and the BMT-1 and LEU-10 side chains are parallel to the ring plane. The next steps along the path are state 24:(0,1,0,0,0,0), where the LEU-6 side chain has also moved into the ring plane, and then state 29:(2,1,0,2,0,0). CsA can be divided into two regions (Leu10–Abu2, and Gly3–Leu9) roughly comprising the two halves of the molecular ring with flexible hinge regions at Gly3 and Leu9.[38] The intermediate state 29 (see Fig. 9) has one region Leu10–Abu2 that has flipped into a backbone conformation similar to that of the bound conformation, while the two remaining monitored hydrophobic residues Leu10 and Bmt1 have remained in the configuration found in the free conformation (see Fig. 7b). The high population in Figure 9 is an indication of the stability of this state, as identified earlier.[38]

Interestingly, no pathway could be observed where one side chain alone flipped to the other ring plane to initiate the transition from the free to the bound CsA conformation. A cooperative change of at least two side chains seems to be required.

Discussion

As demonstrated by the two examples, the PEPCAT method is simple and allows the user to relate conformational states to changes in real conformational properties of the molecule. It is computationally economical to perform, and the results of the conformational trajectory analysis can be displayed in a succinct, visual, and informative manner.

CLASSIFICATION

The classification step provides a powerful and flexible tool kit to identify related structures. Conformational analysis is performed in a relatively small discrete solution space compared to the very large atomic coordinate space. Conformational states can be defined in accordance with the user's requirements based on (generalized) dihedral angles, ϕ–ψ regions or interatomic distances (see Fig. 2). Residue–residue distance measurements (Fig. 2c) for example, could be used to define a conformational state based on a residue contact map. Other classification features can easily be included in future versions of PEPCAT.

The choice of grid size used for the descriptors in the classification scheme play an important part in determining the potential and actual sizes of the database of known states. A finer grid size will imply a larger potential number of known states, and particularly for the analysis of flexible molecules at elevated temperatures, more of these states will be found in the trajectory data files. The exact choice of data grid size used for each descriptor should, therefore, be made with some consideration.

The conformational identity is not necessarily restricted to structures that share the same amino acid sequence, as indicated in the analysis of the family of peptides related to pyro-EEDCK. The ability to specify separate *descriptors* for each different chemical composition and to map them to a shared database of known states is helpful in this process.

COMPARISON

The comparison method returns a value that reflects a generalized distance between conformations, and gives meaningful results even for structures that are not very similar. The simple comparison procedure described here makes three assumptions: (i) that all possible changes in a measurement value have the same level of significance, i.e., a descriptor is either the same or different; (ii) that each descriptor is independent from other descriptors; and (iii) that all descriptors have the same ranking of importance. It is easy to incorporate modifications such as returning a range of values for measurement differences, or scaling a particular descriptor to emphasize a particular feature.

TRAJECTORY ANALYSIS

The PEPCAT trajectory analysis provides a compact representation of the trajectory data file, and extracts information about conformational state connectivity directly from the trajectory. It does not require N^2 comparisons for N frames, and easily allows the incorporation of additional data from new trajectory profiles.

If the dynamics of conformational transitions need to be analyzed, it is important that the conformational state of the system does not undergo more than one change per frame in the trajectory. If the time step with which frames are stored is longer than the time taken for a single transition, multiple conformational changes can occur within one time step, resulting in a distorted conformational transition map. The data obtained from such trajectories can still be processed, for example, in a population analysis (see Fig. 3). Analysis of time sequenced events, however, has been compromised. This problem could also be solved by the inclusion of certain features into the dynamics calculations. A first feature is to record the highest energy value found between two frames that are stored in a trajectory. A second more radical change would be the inclusion of the classification procedure into the molecular dynamics program, so that structures could be analyzed and classified within the dynamics program at the much finer scale of each calculation time step.

OTHER APPLICATIONS

In addition to the analysis of molecular dynamics simulations, our technique could also be applied to other analytical procedures. Conformational search procedures using simulated annealing, Monte Carlo, genetic algorithms, and ligand docking methods could benefit from application of our technique. In the construction and assessment of protein prediction procedures there is often a need to compare a large number of predicted structures against a standard target structure. PEPCAT can be used to measure the structural diversity of the prediction coverage and to assessing the closeness of each predicted structure to the standard structure. QSAR methodologies designed to incorporate

conformational properties could also employ our methodology, because QSAR descriptors could be incorporated in a PEPCAT classification scheme.

Further information, program source code, and access to on-line analysis using the PEPCAT server are available from the PEPCAT web site at: http://www.ludwig.edu.au/pepcat/index.html.

Acknowledgments

We would like to thank Nicos A Nicola (from the Walter and Eliza Hall Institute of Medical Research) for suggesting the studies based on the pyro-EEDCK peptides and ongoing discussion. In addition, we would like to thank Leo Groenen, Tran Trung Tran, David Smith, Robert Jorissen, Nathan Hall, and Jun Zeng for many helpful discussions in the development of PEPCAT.

References

1. Anfinsen, C. B. Science 1973, 181, 223.
2. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D. J Chem Phys 1983, 2, 926.
3. Burgess, A. W.; Scheraga, H. A. Proc Natl Acad Sci USA 1975, 72, 1221.
4. Leach, A. R. Molecular Modelling Principles and Applications; Addison Wesley Longman Ltd: Essex, UK, 1996, chap. 8.
5. Orengo, C. A.; Michie, A. D.; Jones, S.; Jones, D. T.; Swindells, M. B.; Thornton, J. M. Structure 1997, 8, 1093.
6. Murzin, A. G.; Brenner, S. E.; Hubbard, T.; Chothia, C. J. Gen Appl Microbiol 1995, 247, 536.
7. Amadei, A.; Linssen, A. B. M.; Berendsen, H. J. C. Proteins 1993, 4, 412.
8. Vanaalten, D. M. F.; Degroot, B. L.; Finlay, J. B. C.; Berendsen, H. J. C.; Amadei, A. J. Comp Chem 1997, 2, 169.
9. Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. J Comp Chem 1983, 2, 187.
10. Brünger, A. T. X-PLOR Manual; Yale University Press: New Haven, CT, 1992.
11. Verbitsky, G.; Nussinov, R.; Wolfson, H. Proteins Struct Funct Genet 1999, 34, 232.
12. Young, M. M.; Skillman, A. G.; Kuntz, I. D. Proteins Struct Funct Genet 1999, 34, 317.
13. Shenkin, P. S.; McDonald, D. Q. J Comp Chem 1994, 8, 899.
14. Torda, A. E.; Van Gunsteren, W. F. J Comp Chem 1994, 12, 1331.
15. Holm, L.; Sander, C. Mol Biol 1993, 233, 123.
16. Crippen, G. M. J. Comp Phys 1977, 24, 96.
17. Mirny, L.; Domany, E. Proteins Struct Funct Genet 1996, 26, 391.
18. Luger, G. F.; Stubblefield, W. A. Artificial Intelligence and the Design of Expert Systems; Benjamin/Cummings Publishing Company, Inc.: Redwood City, CA, 1989, part II.
19. Moult, J.; Hubbard, T.; Bryant, S. H.; Fidelis, K.; Pedersen, J. T. Proteins Struct Funct Genet Suppl 1997, 1, 2.
20. Lambert, M. H.; Scheraga, H. A. J Comp Chem 1989, 6, 798.
21. Bravi, G.; Gancia, E.; Zaliani, A.; Pegna, M. J Comp Chem 1997, 10, 1295.
22. Kamada, T.; Kawai, S. Inf Process Lett 1989, 1, 7.
23. Tubiana, M.; Carde, P.; Frindel, E. Radiother Oncol 1993, 29, 1.
24. Laerum, O. D.; Frostad, S.; Ton, H. I.; Kamp, D. FEBS Lett 1990, 269, 11.
25. Alisauskas, R. M.; Goldenberg, D. M.; Sharkey, R. M.; Blumenthal, R. D. Int J Cancer 1997, 3, 323.
26. Paukovits, W. R.; Moser, M. H.; Binder, K. A.; Paukovits, J. Blood 1991, 6, 1313.
27. Paukovits, W. R.; Paukovits, J. B.; Moser, M. H.; Konstantinov, S.; Schulte–Hermann, R. Exp Haematol 1998, 26, 851.
28. Maple, J. R.; Dinur, U.; Hagler, A. T. Proc Natl Acad Sci USA 1988, 85, 5350.
29. Wenger, R. M.; France, J.; Bovermann, G.; Wallister, L.; Widmer, A.; Widmer, H. FEBS Lett 1994, 340, 255.
30. Liu, J.; Albers, M. W.; Wandless, T. J.; Luan, S.; Alberg, D. G.; Belshaw, P. J.; Cohen, P.; MacKintosh, C.; Klee, C. B.; Schreiber, S. L. Biochemistry 1992, 31, 3896.
31. Ryffel, B. Pharm Rev 1989, 3, 408.
32. Wüthrich, K.; von Freyberg, B.; Weber, C.; Wider, G.; Traber, R.; Widmer, H.; Braun, W. Science 1991, 254, 953.
33. Allen, F. H.; Kennard, O. Chem Des Auto News 1993, 1, 31.
34. Loosli, H. R.; Kessler, H.; Oschkinat, H.; Weber, H. P.; Petcher, T. J.; Widmer, A. Helv Chim Acta 1985, 68, 682.
35. Pfügl, G.; Kallen, J.; Schirmer, T.; Jansonius, N.; Zurini, G. M.; Walkinshaw, M. D. Nature 1993, 361, 91.
36. Mikol, V.; Kallen, J.; Pfluegl, G.; Walkinshaw, M. D. J Mol Biol 1993, 234, 1119.
37. Sussman, J. L.; Lin, D.; Jiang, J.; Manning, N. O.; Prilusky, J.; Ritter, O.; Abola, E. E. Acta Crystallogr 1998, D54, 1078.
38. O'Donohue, M. F.; Burgess, A. W.; Walkinshaw, M. D.; Treutlein, H. R. Protein Sci 1995, 4, 2191.
39. Mac Kerell, A. D., Jr.; et al. J Phys Chem B 1998, 102, 3586.
40. Pohl, E.; Sheldrick, G. M.; Bolsterli, J. J.; Kallen, J.; Traber, R.; Walkinshaw, M. D. Helv Chim Acta 1996, 6, 1635.